

# Longitudinal Dynamic Functional Regression

Md Nazmul Islam\*

Ana-Maria Staicu†

Eric van Heugten‡

## Abstract

This article develops flexible methodology to study the association between scalar outcomes and functional predictors observed over time, at many instances, in longitudinal studies. We propose a parsimonious modeling framework to study time-varying regression that leads to superior prediction properties and allows to reconstruct full trajectories of the response. The idea is to model the time-varying functional predictors using orthogonal basis functions and expand the time-varying regression coefficient using the same basis. Numerical investigation through simulation studies and data analysis show excellent performance in terms of accurate prediction and efficient computations, when compared with existing alternatives. The methods are inspired and applied to an animal science application, where of interest is to study the association between the feed intake of lactating sows and the minute-by-minute relative humidity throughout the 21st days of their lactation period. R code and an R illustration are provided at <http://www4.stat.ncsu.edu/~staicu/software>.

**Keywords:** Functional data; Functional principal component analysis; Longitudinal study; Longitudinal functional regression; Penalization; Time-varying coefficient model.

## 1 Introduction

Functional regression has attracted a lot of interest in recent years; see Cai et al. (2006); Cardot et al. (1999, 2003); Fan and Zhang (2000); Ivanescu et al. (2015); Morris and Carroll (2006); Müller (2005); Ramsay and Silverman (1997); Reiss and Ogden (2007); Scheipl et al. (2015) to name a few. In this paper we consider longitudinal scalar-on-function regression for scalar outcomes and functional predictors observed repeatedly. This research is motivated by an animal science study of the effect of daily ambient air relative humidity on feed intake of sows during their lactation period. To be specific, a number of sows are observed for several days during their lactation period and for each day, the total daily feed intake, as well as, the minute-by-minute daily relative humidity for the respective day are recorded. Figure 1 illus-

---

\*Department of Statistics, North Carolina State University (Email: [mnislam@ncsu.edu](mailto:mnislam@ncsu.edu))

†Department of Statistics, North Carolina State University (Email: [astaicu@ncsu.edu](mailto:astaicu@ncsu.edu); corresponding author)

‡Department of Animal Science, North Carolina State University (Email: [evheugte@ncsu.edu](mailto:evheugte@ncsu.edu))

trates the data for a randomly chosen sow: daily feed-intake for each lactation day (left panel) and relative humidity profiles (right).

Functional linear model (FLM) for scalar-on-function regression is a popular regression model and assumes that the effect of the functional predictor is captured by the integral of the predictor weighted by a smooth regression coefficient. Three estimation approaches are quite common: both the functional predictor and smooth coefficient are expanded using the empirical eigenbasis of the predictors covariance (Cardot et al. (1999)); both the functional predictor and smooth coefficient are expanded using B-spline basis and penalties are employed to control the smoothness of the parameter function (Ramsay and Silverman (1997)); or a mixture of these approaches, predefined basis function is used to represent the smooth parameter, the empirical eigenbasis of the predictors covariance is used to expand the functional predictors, and in addition penalties are used to control the smoothness of the parameter function (Cardot et al. (2003); Goldsmith et al. (2011)). Extensions of these approaches to accommodate additional covariates or more flexible relationships have been discussed previously by Cardot and Sarda (2008); McLean et al. (2014); Morris and Carroll (2006); Müller and Yao (2012); Ramsay and Silverman (2005).

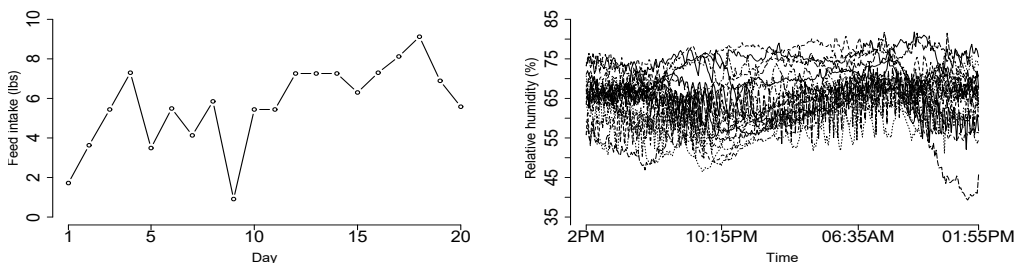


Figure 1: Data for a randomly chosen sow; feed-intake (lbs) and relative humidity (%) profiles from left to right respectively.

The popular scalar-on-function regression model has been extended to analyze the longitudinal data where the scalar response and the functional predictor are observed repeatedly for each subject. Goldsmith et al. (2012) introduced longitudinal penalized functional regression (LPFR), which assumes that the effect of the functional predictor is constant over time. These authors model the time-invariant regression coefficient using the truncated polynomial basis and use a penalized-based approach for estimation. In a similar spirit, Gertheiss et al. (2013) introduced longitudinal functional principal component regression, where the response is regressed onto the functional principal components of the various structures that compose the functional predictors, obtained using longitudinal functional principal component analysis (Greven et al. (2010)). The main limitation of these methods is the assumption that the predictor is invariant over time. Such assumption may be viewed as strong and unrealistic in some situations. For example, in the motivating animal science study, a lactating sow's body can adjust to the prolonged exposure to heat and, thus, the effect of the heat on their feed intake behavior is expected

to change throughout their lactation period. Therefore, assuming a time-varying relationship between the relative humidity and the feed intake is more appropriate.

Recently, Kundu et al. (2016) considered a time-varying functional coefficient in this longitudinal functional model framework. The authors model the functional coefficient as a linear combination of parametric time-dependent functions with smooth time-independent coefficients; they propose to choose the parametric functions based on prior information about the shape of the functional coefficient. However, it is unclear how to obtain such information in our animal science study application. In addition, the methods are limited to Gaussian responses and it is not clear how to extend them to non-Gaussian cases. Our numerical experience with this method indicated that, when applicable, the approach is rather computationally expensive and numerically instable. Furthermore, like LPFR and longitudinal functional principal component regression, this method too does not consider reconstruction of full response trajectory, which is often a major goal in longitudinal studies involving a repeatedly measured response.

In this paper, we propose the longitudinal dynamic functional regression (in short LDFR) for longitudinal scalar-on-function regression that accounts for a time-varying smooth effect of the functional predictors. Unlike the method described in (Kundu et al. (2016)), our modeling framework does not assume known time-structure for the time-varying regression coefficient. The second novelty is that both the proposed modeling and estimation procedure account for the dependence across repeated subject-level measurements of the functional predictor. The third novelty of this approach is that it allows to predict the full response trajectory. The key idea of the proposed approach is to represent both the functional predictors and the unknown regression coefficient using a common time-invariant orthogonal basis functions. The main advantages of our approach are: (i) it enjoys a parsimonious modeling framework; (ii) it does not require any information about the shape of the functional coefficient; (iii) it provides excellent numerical performance in terms of prediction accuracy; (iv) it is computationally very efficient; and (v) it can be easily implemented using well-developed freely available software.

The structure of this paper is as follows. Section 2 introduces the proposed methodology for responses in the exponential family. Section 3 describes the estimation procedure of the parameters of interest. Section 4 details the outcome curve prediction. Numerical assessment of the proposed method is considered in Section 5 in a simulation study and in Section 6, via our motivating application. Finally, Section 7 points out limitations of the proposed approach and discusses feasible extensions.

## 2 Proposed methodology

Let the observed data be  $[t_{ij}, Y_{ij}, \{(W_{ijr}, s_{ijr}) : r = 1, \dots, r_{ij}\}; i = 1, \dots, I, j = 1, 2, \dots, n_i]$ ; where  $i$  indexes the subject,  $j$  indexes the repeated observations,  $Y_{ij}$  is the response measured at time  $t_{ij}$ , and  $W_{ijr}$ 's are the noisy functional covariates observed at points  $s_{ijr}$ 's. It is assumed that  $t_{ij} \in \mathcal{T}$  and  $s_{ijr} \in \mathcal{S}$  for closed and compact sets  $\mathcal{T}$  and  $\mathcal{S}$ , respectively. We assume that for each  $i$  and  $j$ ,  $r_{ij}$  is large, and furthermore that the set  $\{s_{ij1}, \dots, s_{ijr_{ij}}\}$  is a finite grid in  $\mathcal{S}$ . We consider

that  $W_{ijr} = X_{ij}(s_{ijr}) + \epsilon_{ijr}^w$ , where  $X_{ij}(\cdot)$  is the latent functional predictor corresponding to the subject  $i$  and time  $t_{ij}$ , and  $\epsilon_{ijr}^w$  is a measurement error.

Our objective is to develop association models for time-varying responses  $Y_{ij} = Y_i(t_{ij})$  and true smooth time-varying functional covariate  $X_{ij}(\cdot) = X_i(\cdot, t_{ij})$ . Specifically, we consider longitudinal dynamic functional regression models that account for time-varying smooth effect:

$$\begin{aligned} Y_{ij}|X_{ij}(s) &\sim \text{EF}(\mu_{ij}, \eta), \\ g(\mu_{ij}) &= \alpha(t_{ij}) + \int_{\mathcal{S}} X_{ij}(s) \gamma(s, t_{ij}) ds + Z_{b,ij} b_i, \end{aligned} \quad (1)$$

where  $\text{EF}(\mu_{ij}, \eta)$  denotes the exponential family with mean  $\mu_{ij}$  and dispersion parameter  $\eta$  and  $g(\cdot)$  is a known, monotone link function. Here,  $\alpha(\cdot)$  is an unknown intercept function, and  $\gamma(\cdot, t_{ij})$  is an unknown parameter function that quantifies the time-varying effect of  $X_{ij}(\cdot)$  on the conditional mean response of  $Y_{ij}$  at  $t_{ij}$  and is the main object of interest. Also  $b_i$  is a subject specific  $q$ -dimensional vector and  $Z_{b,ij}$  is its associated  $1 \times q$ -dimensional random design matrix. It is assumed that  $b_i$  are independent and identically distributed as  $N_q(0_q, D)$  where  $0_q$  is the  $q$ -dimensional vector of zeros and  $D$  is  $q \times q$  covariance matrix. Both the intercept function  $\alpha(\cdot)$  and the regression coefficient function  $\gamma(\cdot, \cdot)$  are assumed to be smooth. Model (1) has been introduced in Kundu et al. (2016). The key difference is that Kundu et al. (2016) assume that for each  $s$ ,  $\gamma(s, \cdot)$  is a parametric function, such as a polynomial function,  $\gamma(s, t) = \gamma_0(s) + t\gamma_1(s) + t^2\gamma_2(s)$  with unknown functions  $\gamma_0(\cdot)$ ,  $\gamma_1(\cdot)$ , and  $\gamma_2(\cdot)$ . We impose no such limitation and propose a parsimonious modeling framework that results in excellent prediction performance and very competitive computational properties. For convenience, assume  $s_{ijr} = s_r$ .

Our modeling approach consists of two parts. First, we model the longitudinal functional covariates using an orthogonal time-invariant basis expansion that facilitates the description of the dependence of the functional observations within the same subject. Let  $X_{ij}(s) = \sum_{k=1}^{\infty} \xi_{ijk} \phi_k(s)$ , where  $\{\phi_k(\cdot) : k \geq 1\}$  is a time-invariant orthonormal basis in  $L^2(\mathcal{S})$ ; i.e.  $\int_{\mathcal{S}} \phi_k(s) \phi_{k'}(s) ds = 1$  if  $k = k'$  and 0 otherwise. Moreover  $\xi_{ijk}$ 's are the time-varying basis coefficients defined by  $\xi_{ijk} = \int_{\mathcal{S}} X_{ij}(s) \phi_k(s) ds$  and are dependent over  $j$ . Second, we expand the functional regression coefficient  $\gamma(\cdot, t_{ij})$  using the same basis function  $\gamma(s, t_{ij}) = \sum_{k=1}^{\infty} \beta_k(t_{ij}) \phi_k(s)$ , where  $\beta_k(\cdot)$  is an unknown smooth function of time and defined uniquely by  $\beta_k(t) = \int_{\mathcal{S}} \gamma(s, t) \phi_k(s) ds$ . It follows that  $\int_{\mathcal{S}} X_{ij}(s) \gamma(s, t_{ij}) ds = \sum_{k=1}^{\infty} \xi_{ijk} \beta_k(t_{ij})$ , which yields the following more convenient representation of the model (1):  $g(\mu_{ij}) = \alpha(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ijk} \beta_k(t_{ij}) + Z_{b,ij} b_i$ . In practice the infinite summation is truncated at some finite level, say  $K$ ; selecting  $K$  is discussed in the next section. Using the finite truncation  $K$ , model (1) can be approximated by:

$$g(\mu_{ij}) = \alpha(t_{ij}) + \sum_{k=1}^K \xi_{ijk} \beta_k(t_{ij}) + Z_{b,ij} b_i. \quad (2)$$

In this form, it is easy to see that additional vector covariates can be incorporated in a straightforward manner either through a linear or through a smooth dependence.

## 2.1 Selection of the orthogonal basis

There are several approaches to select  $\phi_k(\cdot)$ 's: one approach is using the eigenbasis of some appropriately chosen covariance function; see Park and Staicu (2015). Another approach is using a pre-specified basis similar to Zhou et al. (2008). Here, we consider the former data-driven approach and briefly summarize it below. The approach also takes into account the fact that  $X_{ij}(\cdot)$ 's are observed indirectly through the noisy functional covariates  $W_{ijr} = W_{ij}(s_r)$ .

Recall that the functional covariate is viewed as the sum of two independent processes  $W_{ij}(s) = X_{ij}(s) + \epsilon_{ij}^w(s)$ , where by the abuse of notation we write  $W_{ij}(\cdot) = W_i(\cdot, t_{ij})$  and  $X_{ij}(\cdot) = X_i(\cdot, t_{ij})$ . Furthermore  $\epsilon_{ij}^w(\cdot)$  is a dependent error process at  $t_{ij}$ . Let  $\Sigma(s, s')$  be the marginal covariance function induced by the true signal  $X_i(\cdot, \cdot)$  and defined by  $\Sigma(s, s') = \int_{\mathcal{T}} E[X_i(s, t)X_i(s', t)]h(t)dt$ , where  $h(\cdot)$  is the sampling density of the time points  $t_{ij}$ 's; see Park and Staicu (2015) for justification that this function is a proper covariance function. Furthermore, let the covariance of  $\epsilon_{ij}^w(\cdot)$  be represented as  $\text{cov}(\epsilon_{ij}^w(s), \epsilon_{ij}^w(s')) = \Gamma(s, s') + \sigma_w^2 \mathbb{1}(s = s')$ , where  $\Gamma(\cdot, \cdot)$  is a smooth covariance function, and  $\mathbb{1}(s = s')$  is an indicator function that equals to 1 if  $s = s'$  and 0 otherwise. We denote by  $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$  and observe that this is a proper covariance function (positive semidefinite and symmetric function); let  $\{\phi_k(\cdot), \lambda_k\}_k$  be the eigen-components of  $\Xi(s, s')$ . Then  $X_{ij}(\cdot)$  can be represented by  $X_{ij}(\cdot) = \sum_{k=1}^K \xi_{ijk} \phi_k(\cdot)$  with the basis coefficients as  $\xi_{ijk} = \int_{\mathcal{S}} X_{ij}(s) \phi_k(s) ds$ . Employing the techniques described in Park and Staicu (2015), one can consistently estimate  $\Xi(\cdot, \cdot)$ , and  $\phi_k(\cdot)$ 's as well as  $\xi_{ijk}$ 's; the finite truncation  $K$  is selected using the percentage of variance explained (PVE) of the marginal covariance function  $\Xi(s, s')$ . Selection of the number of eigenfunctions of a covariance in this manner has been commonly used in the literature; see Di et al. (2009); Staicu et al. (2010).

One approach to estimate the basis coefficients  $\xi_{ijk}$  is directly, using the noisy observed functional predictors  $W_{ij}(\cdot)$ . Such an approach captures the dependence of the functional predictors within the same subject but it does not model it. However in order to predict the response trajectory  $Y_i(t)$  for all  $t$  we need to view  $\xi_{ijk} = \xi_{ik}(t_{ij})$  and model the covariance of  $\text{cov}\{\xi_{ijk}, \xi_{ij'k}\} = G_k(t_{ij}, t_{ij'})$ ; we can reconstruct  $\xi_{ik}(t)$  for all  $t$ , using the assumed covariance structure. One option is to consider parametric covariance models such as exponential, or Matérn, or random effects based models. Another option is to use a nonparametric covariance model for  $G_k(\cdot, \cdot)$ ; in the same spirit as Park and Staicu (2015). The spectral decomposition of  $G_k(\cdot, \cdot)$  is  $G_k(t, t') = \sum_{l \geq 1} \eta_{kl} \psi_{kl}(t) \psi_{kl}(t')$ , where  $\{\eta_{kl}, \psi_{kl}(\cdot)\}$  is the pair of eigenvalues and eigenfunctions for  $\eta_{k1} \geq \eta_{k2} \geq \dots > 0$ , and  $\{\psi_{kl}(\cdot)\}_{l \geq 1}$  are mutually orthogonal, and have unit norm in  $L^2(\mathcal{T})$ . Using the truncated Karhunen-Loève (KL) expansion we can represent  $\xi_{ij}(t) = \sum_{l \geq 1} \zeta_{ikl} \psi_{kl}(t)$ , where,  $\zeta_{ikl} = \int_{\mathcal{T}} \xi_{ik}(t) \psi_{kl}(t) dt$  is random, with zero-mean and variance equal to  $\eta_{kl}$ . The infinite summation is truncated to a finite constant. This approach is typically used in functional principal component analysis and we employ it here as well.

## 2.2 Smooth effects modeling in the approximating longitudinal model

The univariate functions in model (2),  $\alpha(t)$  and  $\beta_1(t) \dots \beta_K(t)$  are unknown smooth functions. Assume for now that  $X_{ij}(\cdot)$ 's and furthermore  $\xi_{ijk}$ 's are known. The implied approximating model is known in the longitudinal data literature as a time-varying coefficient model; see Hoover et al. (1998). This model is a special case of the partially linear model investigated by Moyeed and Diggle (1994) and Zeger and Diggle (1994), where only the intercept coefficient is considered to be time-varying while other parameters remain time-invariant. Furthermore, this model reduces to the varying coefficient model described in Hastie and Tibshirani (1993) for independent scalar responses.

We use basis expansions to model the smooth parameter functions in (2). Although the developments focus on the class of functions of truncated linear penalized splines, the results are general and can be used for other types of basis, such as B-splines or Fourier basis, and other types of penalties. Let  $\alpha(t) = \beta_{00} + \beta_{01}t + \sum_{l=1}^{L_0} u_{0l}(t - \kappa_{0l})_+$  and  $\beta_k(t) = \beta_{k0} + \beta_{k1}t + \sum_{l=1}^{L_k} u_{kl}(t - \kappa_{kl})_+$  where  $\beta_{k0}, \beta_{k1}$ 's are unknown fixed parameters,  $u_{kl}$ 's are independent with  $u_{kl} \sim N(0, \sigma_k^2)$  for  $k = 0, 1, \dots, K$ , and  $\kappa_{01}, \dots, \kappa_{0K}$  are knots in  $\mathcal{T}$ ; here  $(x)_+ = \max(0, x)$ . For each  $k = 0, 1, \dots, K$ , the coefficients of the non-polynomial functions are assumed random and their variance,  $\sigma_k^2$  controls the smoothing of the unknown functions,  $\alpha(\cdot)$  or  $\beta(\cdot)$ 's; see Ruppert et al. (2003).

For simplicity we assume  $L_1 = L_1 = \dots = L_K = L$  and  $\kappa_{kl} = \kappa_l$  for all  $k = 0, 1, \dots, K$  and  $l = 1, \dots, L$ . This yields the following mixed effects representation of  $g(\mu_{ij}) = X_{ij}\beta + Z_{ij,0}u_0 + \xi_{ij1}Z_{ij,1}u_1 + \dots + \xi_{ijK}Z_{ij,K}u_K$ , where  $X_{ij} = [1, t_{ij}, \xi_{ij1}, t_{ij}\xi_{ij1}, \dots, \xi_{ijK}, t_{ij}\xi_{ijK}]$  is a  $2(K+1)$ -dimensional row vector,  $\beta = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}, \dots, \beta_{K0}, \beta_{K1})^T$  is the full vector of fixed effects. Also let  $Z_{ij,k}$  is the  $L$ -dimensional row vector of  $(t - \kappa_l)_+$ 's,  $u_k = (u_{k1}, \dots, u_{kL})^T$  is the vector of random effects and  $u_k \sim N(0_L, \sigma_k^2 I_L)$ , where  $I_L$  is the  $L \times L$  identity matrix, for  $k = 0, 1, \dots, K$ . By an abuse of notation let  $\xi_{ij0} = 1$  for all  $i$  and  $j$ ; then we obtain

$$g(\mu_{ij}) = X_{ij}\beta + \sum_{k=0}^K \xi_{ijk}Z_{ij,k}u_k + Z_{b,ij}b_i. \quad (3)$$

A common viewpoint is to choose a value  $L$  that is sufficiently large to capture the complex structure of the parameter functions and then control the smoothness of the regression functions through penalties. This approach is typically used in non-parametric and semi-parametric regression; see for example Eilers and Marx (1996, 2010); Fahrmeir et al. (2004); Fitzmaurice et al. (2008); Ruppert (2002); Ruppert et al. (2003); Wood (2006b). Estimation of the model parameters is obtained by minimizing a penalized criterion and is discussed next.

## 3 Estimation and prediction

In this section we detail the estimation of the model components, which are separated into covariates related components, such as  $\phi_k(\cdot)$  and  $\xi_{ijk}$ 's and response model components, such as  $\beta$ 's and  $u$ 's. Prediction of the response trajectory  $Y_i(\cdot)$  is detailed in Section 4.

### 3.1 Estimation of the covariates related components

We demean the observed functional predictor and denote the demeaned data by  $\widetilde{W}_{ij}(\cdot)$ . Next  $\widetilde{W}_{ij}(\cdot)$ 's is used to estimate the marginal covariance function  $\Xi(s, s')$ . A consistent estimator of  $\Xi(s, s')$  is obtained by the pooled sample covariance defined as  $\widetilde{\Xi}(s_r, s_{r'}) = \{\sum_{i=1}^I \sum_{j=1}^{n_i} \widetilde{W}_{ijr} \widetilde{W}_{ijr'}\} / \sum_{i=1}^I n_i$ . This estimator is not smooth and may be viewed as a raw estimator of  $\Xi(s, s')$ . The diagonal terms of  $\widetilde{\Xi}(s_r, s_{r'})$  is possibly inflated by the variance of white noise. One can ignore the diagonal terms to obtain the smoothed covariance estimator. Another approach is to use the sandwich smoother (Xiao et al., 2013) to smooth the observed predictor and use this for covariance estimation  $\widehat{\Xi}(s, s')$ . We adopt the later approach for our numerical investigation. Consequently, we estimate the eigen-components  $\{\widehat{\phi}_k(\cdot), \widehat{\lambda}_k\}_k$  from the spectral decomposition of  $\widehat{\Xi}(s, s')$  such as  $\widehat{\Xi}(s, s') = \sum_{k=1}^K \widehat{\lambda}_k \widehat{\phi}_k(s) \widehat{\phi}_k(s')$ , where  $\int_{\mathcal{S}} \widehat{\phi}_k(s) \widehat{\phi}_{k'}(s) ds = 1$  if  $k = k'$  and 0 otherwise, and  $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_K \geq 0$ . Note that the truncation value  $K$  is chosen using the procedure described in Staicu et al. (2010). Using numerical integration, the time-varying loadings  $\widetilde{\xi}_{ijk}$ 's are estimated as  $\widetilde{\xi}_{ijk} = \int_{\mathcal{S}} \widetilde{W}_{ij}(s) \widehat{\phi}_k(s) ds$  for  $k = 1, \dots, K$ .

### 3.2 Estimation of the response related components

The estimation of the model parameters  $\alpha(\cdot)$  and  $\beta_k(\cdot)$ 's for  $k = 1, \dots, K$  in (2) entails estimation of  $\beta$ 's and  $u_k$ 's in (3), where  $\xi_{ijk}$  are replaced by  $\widetilde{\xi}_{ijk}$ . For exposition simplicity denote by  $\widetilde{Z}_{ijk} = \widetilde{\xi}_{ijk} Z_{ij,k}$  for all  $i, j$  and  $k$  and denote by  $\widetilde{X}_{ij}$  the vector  $X_{ij}$  in which  $\widetilde{\xi}_{ijk}$ 's are used in place of  $\xi_{ijk}$ 's. It follows that  $g(\mu_{ij}) = \widetilde{X}_{ij} \beta + \sum_{k=0}^K \widetilde{Z}_{ij,k} u_k + Z_{b,ij} b_i$ ; remark that the subject specific effects  $b_i$ 's account for the dependence across repeated observations within the same subject. Then the model for  $Y$  can be approximated by the following generalized mixed effects model

$$\begin{aligned} Y &\sim EF(\mu, \eta) \\ \mu &= \widetilde{X} \beta + \widetilde{Z} u + Z_b b, \\ u &\sim N(0_{L+LK}, \text{diag}\{\sigma_0^2, \sigma_1^2, \dots, \sigma_K^2\} \otimes I_L) \text{ and } b \sim N(0_{I_q}, I_I \otimes D) \\ u \text{ and } b &\text{ are mutually independent} \end{aligned} \quad (4)$$

where  $\mu$  is obtained by columnwise stacking  $(\mu_{i1}, \dots, \mu_{in_i})^T$ ,  $\widetilde{X}$  is obtained by stacking columnwise  $\widetilde{X}_{ij}$  first over  $j$  and then over  $i$ . Here  $u = (u_0^T | u_1^T | \dots | u_K^T)^T$  and  $b = (b_1^T | \dots | b_I^T)^T$ ,  $\widetilde{Z} = (\widetilde{Z}_0 | \widetilde{Z}_1 | \dots | \widetilde{Z}_K)$  with  $\widetilde{Z}_k$  obtained like  $\widetilde{X}$  by stacking columnwise  $Z_{ij,k}$  over  $j$  and  $i$  and  $Z_b = \text{diag}\{Z_{b,1}, \dots, Z_{b,I}\}$  and  $Z_{b,i}$  is obtained by stacking columnwise  $Z_{b,ij}$  over  $j = 1, \dots, n_i$ .

For given values of the covariance parameters,  $\sigma_k^2$ 's and  $D$ , the estimates of  $\beta$ ,  $u_k$ 's and  $b_i$ 's are the same as the minimizers of the following penalized criterion:

$$pl(\beta, u_0, \dots, u_K, b, \eta) = \sum_{i=1}^I \sum_{j=1}^{n_i} \ell(\beta, u_0, \dots, u_K, b, \eta) + b^T (I_I \otimes D^{-1}) b + \lambda_0 \|u_0\|^2 + \dots + \lambda_K \|u_K\|^2, \quad (5)$$

where  $\ell(\beta, u_0, \dots, u_K, b, \eta)$  is the negative twice the log-likelihood function corresponding to the assumed conditional model for  $Y_{ij}$  and  $\lambda_k = 1/\sigma_k^2$  for  $k = 0, 1, \dots, K$ . To estimate the

variance parameters, a Bayesian perspective where the parameters are estimated using the log of a corresponding marginal likelihood, is more appealing. The ideas are described in Wood (2011) and Wood et al. (2016) and rely on using the Laplace approximation to calculate the desired marginal log-likelihood. When the responses are Gaussian, the implied marginal log-likelihood corresponds to the restricted maximum likelihood (REML).

It follows that the estimated parameter functions are  $\widehat{\alpha}(t) = \widehat{\beta}_{00} + \widehat{\beta}_{01}t + \sum_{l=1}^{L_0} \widehat{u}_{0l}(t - \kappa_{0l})_+$  and  $\widehat{\beta}_k(t) = \widehat{\beta}_{k0} + \widehat{\beta}_{k1}t + \sum_{l=1}^{L_k} \widehat{u}_{kl}(t - \kappa_{kl})_+$  for  $k = 1, \dots, K$  and furthermore

$$\widehat{\gamma}(s, t) = \sum_{k=1}^K \widehat{\phi}_k(s) \widehat{\beta}_k(t). \quad (6)$$

Modeling the smoothing of the unknown functions explicitly, as the inverse of a variance component, allows us to clearly describe the corresponding generalized mixed effects model (4). The criterion (5) can be easily modified to account for other bases and associated penalties: the basis function expansions will modify the term  $\widetilde{X}_{ij}\beta + \sum_{k=0}^K \widetilde{Z}_{ij,k}u_k$  that appears in the expression of  $\ell(\beta, u_0, \dots, u_K, b, \eta)$ , while the associated penalties will modify the term  $\sum_{k=0}^K \lambda_k \|u_k\|^2$ . The smoothing parameters  $\lambda$ 's will maintain the same interpretation; see for example Ivanescu et al. (2015); Wood (2006a).

In our simulation experiment we used B-splines and associated penalties that penalize the roughness of the functions using the the second order difference penalty; see Eilers and Marx (1996), Marx and Eilers (1998), Ruppert et al. (2003), and Marx (2010). The changes implied by this choice are illustrated on the generic time-varying coefficient  $\beta_k(t)$ . Denote the B-spline basis by  $\{B_1(\cdot), \dots, B_L(\cdot)\}$ , represent  $\beta_k(t) = \sum_{l=1}^L B_l(t)\beta_{kl}$ ; its associated penalty is  $\lambda_k P_k(\beta_k) = \lambda_k \beta_k^T S^T S \beta_k$ , where  $S$  is the second order differencing matrix defined as:

$$S_{(L-2) \times L} = \begin{pmatrix} 1 & -2 & 1 & \dots & \dots & \dots & 0 \\ 0 & 1 & -2 & \dots & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \dots & \dots & \cdot \\ 0 & 0 & 0 & \dots & \dots & \dots & 1 \end{pmatrix}.$$

As in (5)  $\lambda_k$  will be used to control the amount of the penalization. The penalized criterion corresponding to such choice could also be represented as (5), but after additional reparameterization of the vector  $(\beta_{k1}, \dots, \beta_{kL})$ . To estimate the smoothing parameters  $\lambda_k$ 's, one can use other approaches than REML such as generalized cross validation (GCV) (Golub et al., 1979; Wood, 2006a). We used both REML and GCV to select the smoothing parameters in our numerical investigation and observed competitive efficiency in terms of prediction error, though REML tends to be more computationally intensive. For this purpose, the numerical results presented in Sections 5 are based on GCV. For the data application we considered REML-based estimation; the results using GCV are included in the Supplementary Material, Section B.2.

## 4 Prediction

The responses  $Y_{ij}$  are predicted directly by substituting the estimates of the parameter functions and the predicted random effects into equation (4). However, in order to predict the entire



trajectories of the response  $Y_i(\cdot)$  we need reconstruction of the basis coefficients trajectories  $\xi_{ik}(\cdot)$ . Following the ideas of Park and Staicu (2015), we do this separately for each  $k$ . Consider  $\{(\tilde{\xi}_{ijk}, t_{ij}) : j = 1, \dots, n_i\}_i$ ; we estimate their covariance  $G_k(t, t')$  using functional principal components analysis for sparse sampling design (Yao et al. (2005)). Estimation of the functional principal components and functional principal components scores,  $\widehat{\psi}_{kl}(\cdot)$  and  $\widehat{\zeta}_{ikl}$  respectively allows to estimate the full trajectories  $\widehat{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \widehat{\zeta}_{ikl} \widehat{\psi}_{kl}(t)$  at any time  $t \in \mathcal{T}$ .

Consider the case when the responses are Gaussian. In the case of a random intercept and random slope,  $b_i = (b_{i0}, b_{i1})^T$ , the response trajectories can be predicted using  $E[Y_i(t)|X_i(s, t), b_i]$ :

$$\widehat{Y}_i(t) = \widehat{\alpha}(t) + \sum_{k=1}^K \widehat{\xi}_{ik}(t) \widehat{\beta}_k(t) + \widehat{b}_{i0} + \widehat{b}_{i1}t, \quad (7)$$

where  $\widehat{b}_{i0}$  and  $\widehat{b}_{i1}$  are the predicted subject specific effects.

When the responses are non-Gaussian then the link function is used to predict the response trajectories. Consider for example Bernoulli responses with logit link function and assumed the same random intercept and random slope as above. Then if  $g^{-1}\{\widehat{\alpha}(t) + \sum_{k=1}^K \widehat{\xi}_{ik}(t) \widehat{\beta}_k(t) + \widehat{b}_{i0} + \widehat{b}_{i1}t\} \geq 0.5$  we predict  $\widehat{Y}_i(t) = 1$  and we predict  $\widehat{Y}_i(t) = 0$ , otherwise.

We presented the estimation of the parameter functions  $\alpha(\cdot)$  and  $\beta_k(\cdot)$  based on  $\tilde{\xi}_{ijk}$ ; however our numerical investigation showed slightly improved performance if  $\widehat{\xi}_{ik}(t_{ij})$  are used instead; in fact the numerical results presented in Section 5 are based on using  $\widehat{\xi}_{ik}(t_{ij})$  in the estimation of the parameter functions. The performance of the predicted response trajectories is assessed numerically in the next section.

## 5 Simulation

### 5.1 Description of the settings

We use Monte Carlo simulations to assess the numerical performance of the proposed method (LDFR) and compare it with two other competing models: LPFR (Goldsmith et al., 2012) and LPEER (Kundu et al., 2016). The data  $[t_{ij}, Y_{ij}, \{(W_{ijr}, s_r) : r = 1, \dots, R\} : j = 1, \dots, n_i]_{i=1}^I$  are generated according to the following scenarios:

$$(A) \quad X_i(s_r, t) = \tau(s_r, t) + \sqrt{2}\zeta_{i11}\cos(2\pi t)\cos(2\pi s_r) + \sqrt{2}\zeta_{i12}\sin(2\pi t)\cos(2\pi s_r) + \sqrt{2}\zeta_{i21}\cos(4\pi t)\sin(2\pi s_r) + \sqrt{2}\zeta_{i22}\sin(4\pi t)\sin(2\pi s_r),$$

where  $\tau(s_r, t) = 1 + 2s_r + 3t + 4s_r t$ . Moreover,  $\zeta_{i11}$ ,  $\zeta_{i12}$ ,  $\zeta_{i21}$ , and  $\zeta_{i22}$  are assumed to be mutually independent and identically distributed (IID) such as  $\mathcal{N}(0, 3.5)$ ,  $\mathcal{N}(0, 2)$ ,  $\mathcal{N}(0, 3)$ , and  $\mathcal{N}(0, 1.5)$  respectively. Let  $W_{ijr} = W_i(s_r, t_{ij}) = X_i(s_r, t_{ij}) + \epsilon_{ij}^w(s_r)$ . We define the error term as  $\epsilon_{ij}^w(s_r) = \sqrt{2}\cos(2\pi s_r)\epsilon_{1,ij} + \sqrt{2}\sin(2\pi s_r)\epsilon_{2,ij} + \epsilon_{3,ij}(s_r)$ . Here,  $\epsilon_{1,ij}$ ,  $\epsilon_{2,ij}$ , and  $\epsilon_{3,ij}(s_r)$  are IID such as  $\mathcal{N}(0, \sigma_{\epsilon_1}^2)$ ,  $\mathcal{N}(0, \sigma_{\epsilon_2}^2)$ , and  $\mathcal{N}(0, \sigma_{\epsilon_3}^2)$ ; where,  $\sigma_{\epsilon_1}^2 = 0.3$ ,  $\sigma_{\epsilon_2}^2 = 0.7$ , and  $\sigma_{\epsilon_3}^2$  are calculated using signal-to-noise-ratio (SNR) which is defined as

$$SNR = \frac{\int_{\mathcal{T}} \int_{\mathcal{S}} \text{var}\{W_i(s, t)\} ds dt}{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2 + \sigma_{\epsilon_3}^2} - 1.$$

(B1) *Large noise variance*:  $\sigma_{\epsilon_3}^2 = 9$  (i.e.  $SNR = 0.5$ ).

(B2) *Small noise variance*:  $\sigma_{\epsilon_3}^2 = 1$  (i.e.  $SNR = 2.5$ ).

We consider a dense design for  $s$ :  $\{s_1, \dots, s_R\}$  is taken as a grid of 101 equidistant points in  $[0, 1]$ . We consider two sampling designs for time points  $t_{ij}$ 's corresponding to the number of repeated measurements:

(C1) *Sparse design* when the number of repeated responses per subject is:  $n_i \in \{6, \dots, 10\}$ ,

(C2) *Moderately sparse design* when  $n_i \in \{16, \dots, 20\}$ .

In each case  $\{t_{i1}, \dots, t_{in_i}\}$  are randomly chosen from a dense set of 41 equidistant points; i.e.  $t_{ij} \in [0, 1]$ .

We generate  $Y_{ij}$  in the exponential family as follows:

(D1) Gaussian responses with mean  $\mu_{ij}$  defined by  $\mu_{ij} = \alpha(t_{ij}) + \int X_i(s, t_{ij}) \gamma^\delta(s, t_{ij}) ds$ , where  $\alpha(t) = 7 \sin(3\pi t)$  and consider the following three dependence structures:

(i) *Independent covariance structure*:  $\varepsilon_{ij}$  is distributed as IID  $\mathcal{N}(0, 1.5)$ .

(ii) *Compound symmetric (CS) structure*:  $\varepsilon_{ij} = b_{i0} + e_{ij}$ , where,  $b_{i0}$  and  $e_{ij}$  are distributed as IID  $\mathcal{N}(0, 1)$  and  $\mathcal{N}(0, 0.5)$  respectively, and are mutually independent.

(iii) *Random effect model (REM)*:  $\varepsilon_{ij} = b_{i0} + b_{i1}t_{ij} + e_{ij}$ , where,  $b_{i0}$  and  $b_{i1}$  are distributed as IID  $\mathcal{N}(0, 1)$  and  $\mathcal{N}(0, 0.5)$  respectively with  $cov(b_{i0}, b_{i1}) = 0.1$ . Moreover,  $e_{ij}$  is distributed as IID  $\mathcal{N}(0, 0.3)$ , and is independent from both  $b_{i0}$  and  $b_{i1}$ .

(D2) Binary-valued responses. We consider  $P(Y_{ij} = 1) = \exp(\omega_{ij}) / \{1 + \exp(\omega_{ij})\}$ , where  $\omega_{ij} = \alpha(t_{ij}) + \int_S X_i(s, t_{ij}) \gamma^\delta(s, t_{ij}) ds + b_{i0} + b_{i1}t_{ij}$ , and the subject-specific random effects  $b_{i0}, b_{i1}$  are used to model the dependence across the repeated measurements and are generated as in (iii) above. Our initial choice for  $\alpha(\cdot)$  varies with time; in order to compare our method to LPFR, we change this to  $\alpha(\cdot) = 2$ , since the latter does not seem to accommodate time-varying intercept.

We consider the functional coefficient as  $\gamma^\delta(s, t) = \sqrt{2} \exp(-\delta t) \cos(2\pi s) + \sqrt{2} \delta t \sin(\delta t) \sin(2\pi s)$ . Here,  $\delta$  controls the departure from a time-invariant effect: when  $\delta = 0$  then  $\gamma^\delta(s, t) = \sqrt{2} \cos(2\pi s)$ , and for  $\delta \neq 0$ ,  $\gamma^\delta(s, t)$  varies with time  $t$ . We investigate the cases  $\delta = \{0, 1, 2, 5\}$ .

We also examine the performance for varying sample sizes  $I = \{100, 200, 300\}$ . For implementation, we briefly describe the steps here: (1) Estimate the bivariate mean function  $\hat{\tau}(\cdot, \cdot)$  using the fast bivariate smoother based on tensor product of cubic splines with 35 knots in each direction and second order difference penalty (Xiao et al., 2013) and by selecting the smoothing parameter using GCV. (2) Demean the observed predictor and estimate the smoothed marginal covariance using the sandwich bivariate smoother as described earlier. (3) Perform eigenanalysis of the estimated smooth covariance and obtain the pairs of eigenvalues and eigenfunctions

$\{\widehat{\lambda}_k, \widehat{\phi}_k(\cdot)\}_k^K$ ; here  $K$  is chosen using 95% PVE value. (4) Estimate the time-varying loadings  $\widehat{\xi}_{ijk} = \int_{\mathcal{S}} \{W(s, t_{ij}) - \widehat{\tau}(s, t_{ij})\} \widehat{\phi}_k(s) ds$  using numerical integration. (5) For each  $k$ , consider  $\{\widehat{\xi}_{ijk}, t_{ij} : j = 1, \dots, n_i\}_i$  and assume a nonparametric covariance structure for the dependence of  $\widehat{\xi}_{ijk}$  across  $j$  as described in Section 3.1. Furthermore, estimate the eigen-components  $\{\widehat{\eta}_{kl}, \widehat{\psi}(\cdot)_{kl}\}_{l=1}^{L_k}$  where  $\eta_{kl}$  are the non-decreasing non-negative eigenvalues;  $L_k$  is chosen by 95% PVE value. Estimate the time-varying loadings for any time  $t \in \mathcal{T}$ ; i.e.  $\widehat{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \widehat{\xi}_{ikl} \widehat{\psi}_{kl}(t)$ . (6) Use  $\widehat{\xi}_{ijk} = \widehat{\xi}_{ik}(t_{ij})$  in (3) and fit the approximated generalized mixed model with penalties: use an identity link for Gaussian responses and logit link for binary responses, by assuming independence between random effects.

For LPFR the time-invariant regression coefficient  $\gamma(s)$  is modeled using the truncated linear basis with 30 functions (default choice) and the smoothing parameter is estimated by GCV; the model is fitted using the function `lpfr` in the R package `refund` (Huang et al. (2015)). For LPEER, the time-varying coefficient  $\gamma(s, t)$  is represented using a polynomial basis in time  $t$ , with coefficients that are smooth functions in  $s$  and which are estimated using a penalized criterion with a ridge penalty; the degree of the polynomial basis is selected using the Akaike information criterion (AIC) (Akaike (1974)) and the smoothing parameters of the smooth terms are selected using REML. The model is fitted using the function `lpeer` in the R package `refund`.

To assess the performance of the method, we divide each simulated dataset of into a training and test set. Denote by  $I$  the number of subjects and by  $n_i$  the number of instances at which the subject is observed. The test set is formed as follows: for each subject  $i$  in the dataset, we randomly select five instances without replacement and include the corresponding information  $[t_{ij}, Y_{ij}, \{(W_{ijr}, s_r) : r = 1, \dots, R\}]$  in the test data. The remaining observations for each subject are included in the training set. We fit the model using the training data; then we predict both the training data (IN) and test data (OUT) using the estimates obtained from the fit of training data. To evaluate the performance of the competing models for normal responses, we compute the root-mean-prediction-error (RMPE) for training (IN<sub>PE</sub>) and test set (OUT<sub>PE</sub>); i.e.  $\text{IN}_{PE} = \sqrt{\sum_{i=1}^I \{\sum_{j=1}^{n_i-5} (Y_{ij}^{\text{train}} - \widehat{Y}_{ij}^{\text{train}})^2 / (n_i - 5)\} / I}$  and  $\text{OUT}_{PE} = \sqrt{\sum_{i=1}^I \{\sum_{j=1}^5 (Y_{ij}^{\text{test}} - \widehat{Y}_{ij}^{\text{test}})^2 / 5\} / I}$ . For binary-valued responses, we assess the numerical performance in estimating the linear predictor trajectory  $g(\cdot)$ , and with respect to sensitivity or true positive rate (TPR); where TPR is defined as the proportion of successes ( $\widehat{Y}_{ij} = 1$ ) that are correctly identified. For the prediction of entire trajectory, we report the *RMSE* of  $\widehat{Y}_i(\cdot)$  which is evaluated by root mean square error, defined as  $\text{RMSE} = \sqrt{1/I \sum_{i=1}^I [1/n \sum_{j=1}^n \{Y_i(t_j) - \widehat{Y}_i(t_j)\}^2]}$ , where  $\{t_1, \dots, t_n\}$  is an equally spaced grid of 41 number points in  $[0, 1]$ ; the calculation of RMSE is based on the entire data. Results are based on 1000 independent samples for each combination of the simulation settings. Note that, for numerical investigation, we use Intel(R) Core(TM) i7-4770, 3.40 GHz processor with 32.0 GB RAM in 64-bit operating system.

## 5.2 Prediction performance assessment and comparison

First, we consider Gaussian responses ( $D1$ ) and compare our longitudinal dynamic functional regression method (LDFR) with LPFR. Table 1 displays the median of IN-and-OUT sample prediction errors for different  $\delta$  values for 1000 simulations and their respective interquartile ranges (IQR) expressed in the parenthesis. The results correspond to data generated using CS dependence structure ( $D1ii$ ) and fitted by assuming a model with  $CS$  type covariance structure. We observe that prediction errors (i.e. IN-and-OUT) for both methods are similar for  $\delta = 0$ . However, as  $|\delta| > 0$  the functional coefficient is time-dependent and the results show that the proposed approach greatly improves over LPFR. For example, when  $\delta = 5$ , our method may improve prediction accuracy by more than 40% over LPFR. Furthermore, the numerical study shows that the accuracy of our method increases with the number of repeated measurements per subject; this is expected as in this case, the estimation of the within subject covariance improves. In the Supplementary Material, Section A.1 we investigate mild misspecification of the dependence structure and observe similar findings.

Next, we compare the performance of LDFR with LPEER. Because of the heavy computational burden of the latter approach, we limit our investigation to 100 Monte Carlo samples per setting; see Table 2. Here, we fit the competing model without assuming prior knowledge about the structure of the bivariate regression coefficient  $\gamma$ . Table 2 illustrates the performance when data are generated from  $CS$  type covariance structure ( $D1ii$ ). We fit the model by using a random subject intercept model (correct covariance model). When  $|\delta| > 0$  the departure of  $\gamma(s, \cdot)$  from a time-invariant coefficient is stronger. The numerical results show improved performance for our method as  $|\delta| > 0$ . When  $\delta = 1$ , LDFR and LPEER show similar prediction accuracy. However, LDFR is computationally much faster than LPEER; see the right-most columns of Table 2. Furthermore, for  $\delta = 5$ , LDFR outweighs LPEER in nearly all the cases considered; this may be due to the fact that LPEER is designed to exploit the prior knowledge about the shape of  $\gamma(\cdot, \cdot)$ , while we do not use this advantage. In addition, LPEER method is numerically instable for the case when functional covariate evolves heavily over time.

The accuracy in predicting the entire trajectory using the proposed LDFR method is illustrated in Figure 2. The results are RMSEs based on 1000 simulations, for the case of Gaussian responses that are correlated using CS structure ( $D2ii$ ) and are observed in each of the two sampling designs considered: sparse and moderately sparse. As expected, the accuracy improves as the number of repeated measurements per subject increases, as well as when the sample size increases, with the gain being higher for the former situation than for the latter. We also note that the accuracy decreases as  $\delta$  departs from zero, essentially as the regression coefficient departs from a time-invariant function. This finding is in agreement with the results shown in Tables 1 and 2, which compare the prediction performance of our approach with the existing alternatives. As the magnitude of  $\delta$  increases, the difficulty of the problem increases and the prediction accuracy for all the methods suffers; nevertheless the proposed approach yields superior results.

We consider binary responses with logit link ( $D2$ ) and evaluate the prediction accuracy of the proposed method with LPFR, which is the only available appropriate approach. We fit the model with subject-specific random intercept, while data are generated assuming both random intercept and slope (model misspecification). Table 3 shows the estimation error of the linear predictors  $\widehat{g}(\mu_{ij})$  for both sparse and moderately sparse longitudinal design when the functional covariates are observed with large noise. The findings are as expected: when  $\delta = 0$ , the numerical performance of both LDFR and LPFR is similar but as  $\delta$  departs from zero, say for  $\delta = 5$ , the prediction accuracy of LDFR improves substantially. As the magnitude of  $\delta$  increases the results also show an improved performance with respect to the true positive rate in predicting responses using LDFR versus LPFR. Additional simulation results for both Gaussian response and binary response cases are included in the Supplementary Material, Section A.

Table 1: Gaussian responses with CS dependence structure ( $D1ii$ ), when the longitudinal design is sparse ( $C1$ ) and moderately sparse (mod sparse,  $C2$ ); the functional covariates are observed with low noise variance ( $B2$ ). Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

		$\delta = 0$				$\delta = 2$				$\delta = 5$			
		$IN_{PE}$		$OUT_{PE}$		$IN_{PE}$		$OUT_{PE}$		$IN_{PE}$		$OUT_{PE}$	
		LDFR	LPFR	LDFR	LPFR	LDFR	LPFR	LDFR	LPFR	LDFR	LPFR	LDFR	LPFR
I = 100	sparse	0.74	0.86	0.93	0.97	0.83	1.35	1.05	1.51	1.21	3.29	1.63	3.52
		(0.04)	(0.04)	(0.09)	(0.05)	(0.05)	(0.08)	(0.10)	(0.10)	(0.11)	(0.31)	(0.24)	(0.34)
	mod sparse	0.72	0.87	0.78	0.92	0.77	1.39	0.85	1.47	0.98	3.36	1.10	3.49
		(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.07)	(0.05)	(0.09)	(0.07)	(0.26)	(0.13)	(0.32)
I = 200	sparse	0.74	0.85	0.90	0.96	0.82	1.36	1.01	1.51	1.18	3.33	1.51	3.54
		(0.03)	(0.03)	(0.08)	(0.03)	(0.03)	(0.06)	(0.07)	(0.06)	(0.08)	(0.23)	(0.16)	(0.24)
	mod sparse	0.72	0.87	0.78	0.92	0.76	1.39	0.83	1.47	0.95	3.38	1.04	3.50
		(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.05)	(0.04)	(0.07)	(0.05)	(0.20)	(0.08)	(0.27)
I = 300	sparse	0.74	0.85	0.89	0.96	0.82	1.36	0.99	1.51	1.17	3.33	1.45	3.53
		(0.03)	(0.02)	(0.08)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.07)	(0.18)	(0.12)	(0.21)
	mod sparse	0.72	0.87	0.78	0.92	0.76	1.39	0.82	1.46	0.94	3.38	1.03	3.49
		(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.06)	(0.03)	(0.16)	(0.06)	(0.20)

Table 2: Gaussian responses with CS dependence structure ( $D1ii$ ), when the longitudinal design is sparse ( $C1$ ) for  $B1$  and  $B2$ . Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported.

		$\delta = 1$						$\delta = 5$					
		$IN_{PE}$		$OUT_{PE}$		Run time ('sec)		$IN_{PE}$		$OUT_{PE}$		Run time ('sec)	
		LDFR	LPEER	LDFR	LPEER	LDFR	LPEER	LDFR	LPEER	LDFR	LPEER	LDFR	LPEER
I = 100	SNR 0.5	0.72	0.75	0.88	0.96	3.59	111.91	1.26	1.58	1.69	2.40	3.68	472.97
		(0.04)	(0.03)	(0.07)	(0.04)			(0.11)	(0.12)	(0.22)	(0.24)		
	SNR 2.5	0.72	0.76	0.88	0.92	3.31	90.97	1.19	1.41	1.62	2.23	3.59	532.92
		(0.03)	(0.03)	(0.06)	(0.04)			(0.12)	(0.13)	(0.24)	(0.20)		
I = 300	SNR 0.5	0.72	0.79	0.87	0.94	15.75	299.43	1.24	1.79	1.55	2.23	11.88	656.11
		(0.02)	(0.02)	(0.04)	(0.03)			(0.06)	(0.08)	(0.13)	(0.12)		
	SNR 2.5	0.71	0.78	0.85	0.90	13.49	268.54	1.17	1.64	1.47	2.05	14.14	879.49
		(0.02)	(0.02)	(0.04)	(0.02)			(0.06)	(0.10)	(0.11)	(0.09)		

Figure 2: Gaussian responses with CS dependence structure (*D1ii*), when the longitudinal design is sparse (top) and moderately sparse (bottom); the functional covariates are observed with high noise variance (*B1*). Fitted model assumes CS covariance structure. Reported is RMSE based on 1000 simulations.

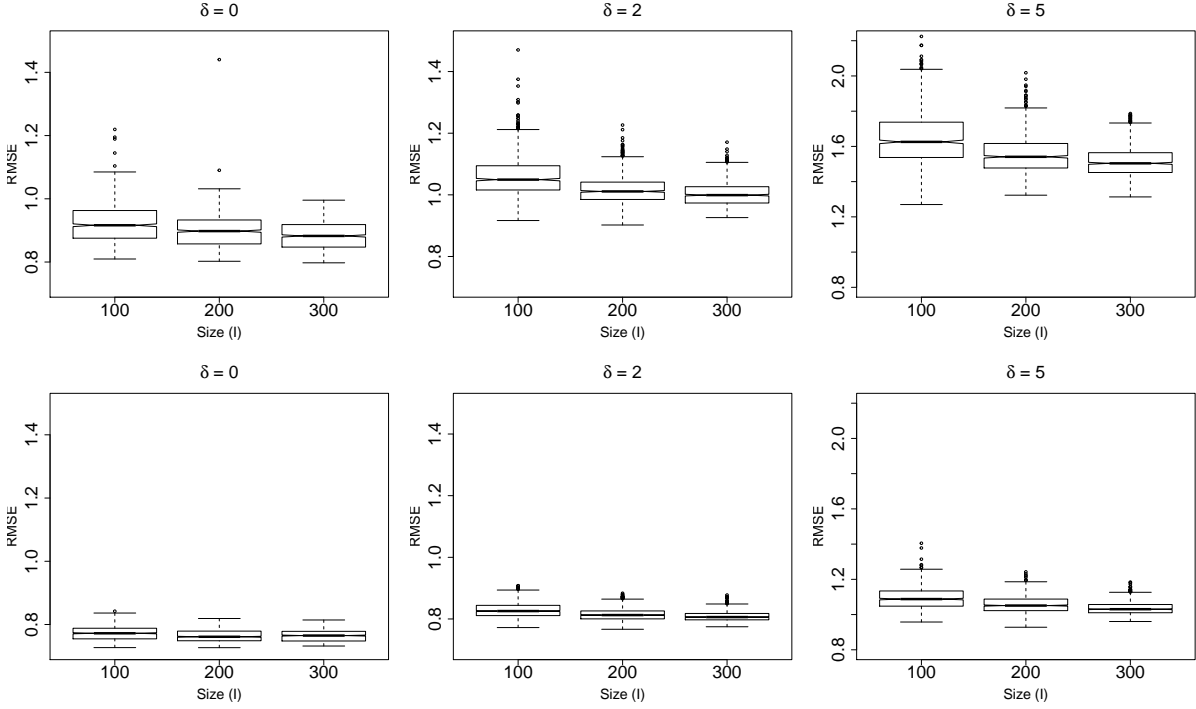


Table 3: Binary responses (*D2*) when the longitudinal design is sparse/moderately sparse (mod sparse); the functional covariates are observed with high noise variance (*B1*). Fitted model assumes subject-specific random intercept. We report the median of prediction errors of the linear predictor trajectories and, in parenthesis, the median of true-positive-rates.

		$\delta = 0$				$\delta = 5$			
		<i>IN<sub>PE</sub></i>		<i>OUT<sub>PE</sub></i>		<i>IN<sub>PE</sub></i>		<i>OUT<sub>PE</sub></i>	
		LDPR	LPFR	LDPR	LPFR	LDPR	LPFR	LDPR	LPFR
I = 100	sparse	1.27	1.30	1.32	1.30	1.82	2.88	2.01	2.90
		(0.96)	(0.96)	(0.96)	(0.96)	(0.91)	(0.87)	(0.90)	(0.86)
	mod sparse	1.19	1.29	1.20	1.29	1.50	2.87	1.55	2.87
		(0.96)	(0.96)	(0.96)	(0.96)	(0.92)	(0.86)	(0.91)	(0.86)
I = 200	sparse	1.23	1.29	1.27	1.29	1.75	2.89	1.90	2.90
		(0.96)	(0.96)	(0.96)	(0.96)	(0.92)	(0.87)	(0.90)	(0.86)
	mod sparse	1.17	1.28	1.18	1.29	1.45	2.87	1.47	2.88
		(0.96)	(0.96)	(0.96)	(0.96)	(0.92)	(0.86)	(0.92)	(0.86)
I = 300	sparse	1.21	1.29	1.25	1.29	1.74	2.89	1.84	2.89
		(0.96)	(0.96)	(0.96)	(0.96)	(0.92)	(0.87)	(0.91)	(0.86)
	mod sparse	1.17	1.28	1.17	1.28	1.41	2.88	1.43	2.88
		(0.96)	(0.96)	(0.96)	(0.96)	(0.92)	(0.86)	(0.92)	(0.86)

## 6 Data application

Our motivating application is a lactating sow study where the primary objective is to investigate the effect of thermal environment (i.e. temperature ( $T$ ) and humidity ( $H$ )) on the feed-intake of the lactating sows. This study is very important for several reasons: (1) ambient temperatures above the evaporative critical temperature decreases the amount of food-intake which, as a result, deteriorates the reproductive performance and hinders the growth rate of piglets of lactating sows (Black et al. (1993)). (2) Also, poor feed-intake of the lactating sows leads to increased body weight loss during lactation and reduced milk yield, and is further associated with compromised weight gain of their litter (Johnston et al. (1999), Renaudeau and Noblet (2001), Renaudeau and Noblet (2001)). (3) Thirdly, heat-stress results in a reduction of farrowing rate (the percentage of sows that become pregnant and farrow a litter of piglets) and total number of pigs born in sows (Bloemhof et al. (2013)) which in turn has a negative effect on the total production of pork meat per year. (4) Fourth, pigs from sows raised in an unfavorable thermal environment will be fatter than the ones reared in favorable cooler environments and this fact makes pork meat fattier (Baumgard (2015)). (5) Fifth, because of heat stress associated with hot climactic thermal environment, the swine industry in US incurs a total estimated loss worth of \$300 million per year on average (St-Pierre et al. (2003)). Therefore, insight into how the feeding behavior changes over time due to the prolonged exposure to a hot environment will assist in proposing more economical and efficient feeding strategies for lactating sows.

The experimental study was carried during July to October in 2013 in a 2,600-sow commercial research unit in Oklahoma (Rosero et al. (2016)) and involved 480 PIC Camborough sows. The sows were kept in the farrowing facility where they gave birth to piglets. Depending on the number of previous pregnancies (parity levels), sows were classified into younger (parity equal to zero or one) or older (parity equal to two or higher). The sows are brought to the farrowing crates when they are approximately five days before they are due to give birth. They arrive in groups; the study involves 21 groups of about 21-23 sows. Sows are observed during their 20-21 day lactation period and their respective food-consumption is monitored. Each sow is provided food individually with a computerized feeding system (Howema, Big Dutchman, Germany). The amount of food-offered ( $FO$ ) was recorded at 2.00PM on each day and feed-refusal ( $FR$ ) was measured the following day prior to any subsequent food addition; feed-intake ( $FI$ ) was calculated as  $FI = FO - FR$  in lbs. Minute-by-minute information about the ambient air temperature ( $^{\circ}C$ ) and relative humidity (%) of the farrowing facility were recorded by data loggers (LogTag, MicroDAQ Ltd., Contoocook, NH). The experimenters removed information of five sows due to unreliable measurements and thus we had available information for 475 sows. The facility ambient temperature was controlled by a ventilation system; the barns have cool cells that pull fresh air through wet corrugated material to provide further cooling of air. There are some missing observations for humidity and temperature profiles due to machine failure which qualifies the pattern of missingness as missing completely at random. Preliminary investigation confirms that the ambient temperature (dry-bulb temperature) and relative humidity

are negatively correlated (see also Lawrence (2005)). For this reason, we focus on the relative humidity in this paper solely; our objective is to study the effect of the relative humidity on the feed intake of sows.

We use  $i$  to index the sows,  $j$  to index the repeated instances for the same sow, and  $t_{ij}$  to denote the lactation day of the  $i$ th sow, corresponding to the  $j$ th instance at which the sow is observed; for many sows we have  $t_{ij} = j$  for  $j = 1, \dots, 21$ , but this is not the general case. Furthermore denote  $g_i$  to index the group of the  $i$ th sow;  $g_i = 1, \dots, 21$ . Typically, the sows within the same group do not give birth in the same day, and thus to study the effect of the ambient humidity during the sows lactation period, it is important to recognize the dependence on the lactation day. Denote by  $nRHum_{ij}(\cdot) = nRHum_i(\cdot, t_{ij})$  the relative humidity daily profile observed for the  $t_{ij}$  lactation day of the  $i$ th sow; the measurements typically include noise, hence the prefix “n”. Later we use notation  $RHum_{ij}(\cdot)$  for the true relative humidity profile corresponding to  $nRHum_{ij}(\cdot)$ . The right panel of Figure 1 shows the relative humidity daily profiles corresponding to all the first 21st days at which the particular random sow was observed. Let  $FI_{ij}$  be the  $FI$  of the  $i$ th sow at its  $j$ th repeated occasion at lactation day  $t_{ij}$ .

We assume that the relationship between the feed intake and the relative humidity is described by the LDFR model:

$$FI_{ij} = \beta_{p_i}(t_{ij}) + \int RHum_{ij}(s)\gamma(s, t_{ij})ds + b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)}t_{ij} + \varepsilon_{ij}, \quad (8)$$

where  $\beta_{p_i}(\cdot)$  is the mean feed intake for group  $p_i$  where  $p_i = 0$  (young sows) and  $p_i = 1$  (old sows) and  $\gamma(s, \cdot)$  quantifies the time-varying effect of the relative humidity on  $FI$ ; the integral reflects the time when the measurements were collected, from 2PM to 1:59PM the following day. The term  $b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)}t_{ij}$  models the dependence of the responses within the same sow as well as of the responses of the sows who are in the same group. Thus  $b_{g_i}$  denotes a group random specific effect and  $b_{0i(g_i)}$  and  $b_{1i(g_i)}$  are sow within the group specific intercept and slope. We assume that  $b_{g_i} \sim N(0, \sigma_g^2)$ ,  $b_{0i(g_i)} \sim N(0, \sigma_0^2)$ , and  $b_{1i(g_i)} \sim N(0, \sigma_1^2)$  are all mutually independent. Finally it is assumed that the measurement errors  $\varepsilon_{ij}$ ’s are independent and distributed as  $N(0, \sigma_e^2)$ .

The steps for fitting the model (8) are as follows. (1) First, estimate the bivariate mean function  $\tau(\cdot, \cdot)$  of the relative humidity profiles  $RHum_i(s, t_{ij})$ . We employ the fast bivariate spline smoothing (Xiao et al. (2013)) using cubic splines with 35 knots in  $s$ -direction and 19 knots in  $t$ -direction; second order difference penalty is used to control the smoothness of the fit and the tuning parameters are estimated by REML. (2) Second, demean the humidity profiles for each observed day  $t_{ij}$  and use the techniques introduced by Park and Staicu (2015) to model the dynamics in longitudinal functional observations; this approach involves smoothing the pooled sample covariance using the fast bivariate spline smoothing technique. (3) Third, obtain the empirical eigencomponents  $\{\widehat{\lambda}_k, \widehat{\phi}_k(\cdot)\}_{k=1}^K$  using the eigen decomposition of the smoothed pooled covariance. The finite truncation  $K$  is selected using 95% PVE value. (4) Fourth, estimate  $\widetilde{\xi}_{ijk}$ ’s by numerical integration, as  $\widetilde{\xi}_{ijk} = \int_S \{nRHum(s, t_{ij}) - \widehat{\tau}(s, t_{ij})\} \widehat{\phi}_k(s) ds$  (5) Fifth, assume a nonparametric covariance model for  $\widetilde{\xi}_{ijk}$ ’s; use functional principal component analysis



technique for sparse data to estimate the covariance function for each  $k$ . Next estimate the eigen-components  $\{\widehat{\eta}_{kl}, \widehat{\psi}(\cdot)_{kl}\}_{l=1}^{L_k}$  where  $\widehat{\eta}_{kl}$  are the non-negative eigenvalues;  $L_k$  is chosen by using 95% PVE. Estimate the time-varying loadings for any time  $t \in \mathcal{T}$ ; i.e.  $\widehat{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \widehat{\zeta}_{ikl} \widehat{\psi}_{kl}(t)$ . (6) Sixth use  $\widehat{\xi}_{ijk} = \widehat{\xi}_{ik}(t_{ij})$ 's in the penalized likelihood function corresponding to model (8) and fit the approximated generalized mixed model with penalties; estimate each parameter function  $\beta_k(\cdot)$  using 10 cubic B-splines basis functions with second order difference penalty. Select the smoothing parameters by REML.

## 6.1 Fit assessment

We first assess the prediction performance of the proposed method and compare it with the available competitors, LPFR and LPEER. In this regard we split the data into a train set, which is used to build the model and a test set on which the prediction performance is evaluated. We consider two ways of forming the test data: (a) randomly select 350 sows and include only the measurements corresponding to their last 10 lactation days; and (b) take all the 475 sows and include only the measurements corresponding to about 20% of their lactation days that are selected at random. The remaining data form the training set. Approach (a) involves data on fewer sows than approach (b). At the same time, the approach (a) assesses the performance of prediction at “future” lactation days, while the approach (b) evaluates the prediction performance at random lactation days within the 1 to 21 days at which the sows are observed.

For completeness we describe the implementation of the competitive approaches. LPFR is fitted by modeling the time-invariant regression coefficient using 30 truncated linear splines basis functions and the tuning parameters are estimated by REML. The covariance structure is specified as in model (8) and the model is fitted using `lpfr()` function available in `refund` package (Huang et al. (2015)). For LPEER, we consider polynomial functions of time of degree  $d = 0, 1, \dots, 4$ ,  $\gamma(s, t) = \gamma_1(s) + t\gamma_1(s) + \dots + t^d\gamma_d(s)$  and select the optimal  $d$  by AIC as described earlier; the model is fitted using `lpeer` in the same package. The covariance structure is specified using a subject-specific random intercept and random group effect; it is not clear how to modify the existing code to accommodate a subject-specific slope effect.

Table 4 reports the results. The findings show that LDFR and LPEER perform relatively similar in terms of in-sample and out-of-sample prediction in the two situations considered. Nevertheless our methodology yields to computations that are an order of magnitude faster and a better fitting criteria as assessed by AIC. In contrast, the prediction results with LPFR are inferior, and they indicate a lack of appropriateness of a time-invariant effect model for our lactating sow application, based on the observations from our simulation investigation.

Table 4: Prediction accuracy, computing time (in seconds), and AIC for the proposed LDFR and the competitive approaches LPFR and LPEER for the sows application.

	<i>IN<sub>PE</sub></i>			<i>OUT<sub>PE</sub></i>			AIC			Computing time		
	LDFR	LPEER	LPFR	LDFR	LPEER	LPFR	LDFR	LPEER	LPFR	LDFR	LPEER	LPFR
(a)	<b>1.21</b>	1.28	1.35	<b>1.66</b>	1.61	1.85	<b>18886</b>	19872	19995	<b>583.00</b>	5476.05	2492.14
(b)	<b>1.32</b>	1.35	1.43	<b>1.41</b>	1.46	1.52	<b>25488</b>	26490	26562	<b>592.10</b>	4842.70	2475.95

## 6.2 Model components estimation and full response trajectory prediction

Next, we fit model (8) using the entire data. Figure 3 illustrates the estimates of the model parameter functions. Specifically the left panel depicts the estimated mean feed intake for old sows (solid line),  $\widehat{\beta}_1(t)$ , and for young sows (dashed line),  $\widehat{\beta}_0(t)$ . It appears that all the sows eat about the same amount of food in the first couple of days of their lactation period, but shortly afterwards the older sows start eating more than their younger relatives. By the fourth lactation day the older sows have an advantage of feed intake of up to 1-1.5 lbs per day and they maintain this advantage for the remaining duration of the lactation period.

Figure 3, in the right panel, shows the estimated regression coefficient  $\widehat{\gamma}(s, t)$ , which quantifies the minute-by-minute ( $s$ ) effect of the relative humidity on the feed intake during the first 21st lactation days ( $t$ ). The interpretation of this estimated effect uses the fact that the relative humidity profiles are centered to have zero mean function. The results indicate a different effect of the humidity throughout the lactation period. Specifically, in the first week, it appears that there are two windows during the day, a two-three hours window in the morning (8:45AM to about 12PM) and a three-hours window in the late evening/night when the effect is positive, with stronger effect in the former morning window. This means that values of relative humidity above the mean level during these window times lead to an increase in the feed intake for the respective day. Later in the lactation period, it seems to be only during the night, between 0AM to 10AM, when humidity has a positive effect on the sows feed intake. This indicates that later in their lactating period, during the night the sows prefer a warmer environment, which seems to be the exactly opposite from their night preference in their first week. In the first lactating days, humidity has a negative effect on the sows' feed intake during the night 0AM to about 5AM. Starting with their second week for about a week, humidity has a negative effect during late morning - early afternoon (from about 10AM to about 2PM) and during late evenings from about 8PM to 11PM. In their last lactation week, this negative effect of humidity seems to expand for the entire duration of the daylight, from about 10AM till 8-9PM. These findings have the potential to help ensuring an improved environment for a better feed intake of lactating sows. Section B.1 of the Supplementary Material includes additional results for the data analysis, while Section B.2 provides prediction results for LDFR approach when a covariance model based on random group and subject random intercept is used.

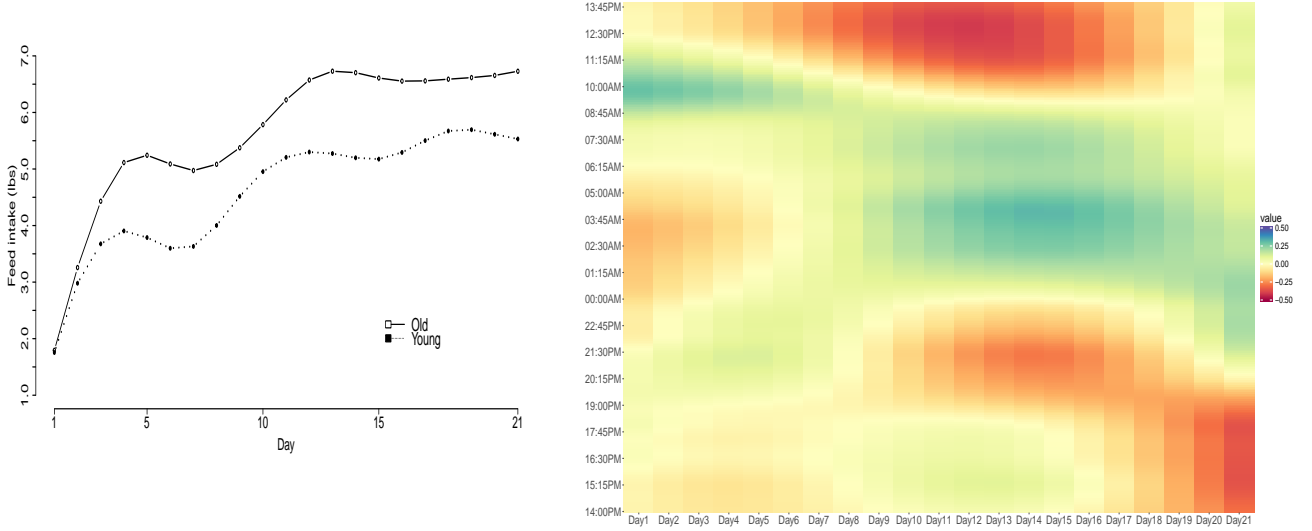


Figure 3: Parameter estimates in the lactation sow application. Left panel depicts the estimated intercept function for the old (solid line) and young (dashed) sows. Right panel shows the estimated regression coefficient  $\gamma(\cdot, t)$  for each lactation day  $t$ .

The estimated standard deviation for the subject random intercept and the random slope are similar,  $\hat{\sigma}_0 = 0.71; \hat{\sigma}_1 = 0.79$ , and the estimated standard deviation of the group effect is much smaller  $\hat{\sigma}_g = 0.20$ . We use these estimates to predict the sow specific intercept and slope effects and the group effects. Figure 4 shows the predicted full trajectories of the feed intake for two young sows (left and middle panels) and one old sow (rightmost) selected at random from different groups. The predicted trajectories are obtained using the expression  $\hat{F}I_i(t) = \hat{\beta}_{p_i}(t) + \int \hat{S}mRHum_i(s, t)\hat{\gamma}(s, t)ds + \hat{b}_{g_i} + \hat{b}_{0i(g_i)} + \hat{b}_{1i(g_i)}t$  for every day  $t = 1, \dots, 21$ , where  $\hat{S}mRHum_i(\cdot, t)$  are the smooth and demeaned relative humidity profiles observed in relation to sow  $i$  and  $\hat{b}_{g_i}, \hat{b}_{0i(g_i)}$  and  $\hat{b}_{1i(g_i)}$  are predicted effects.

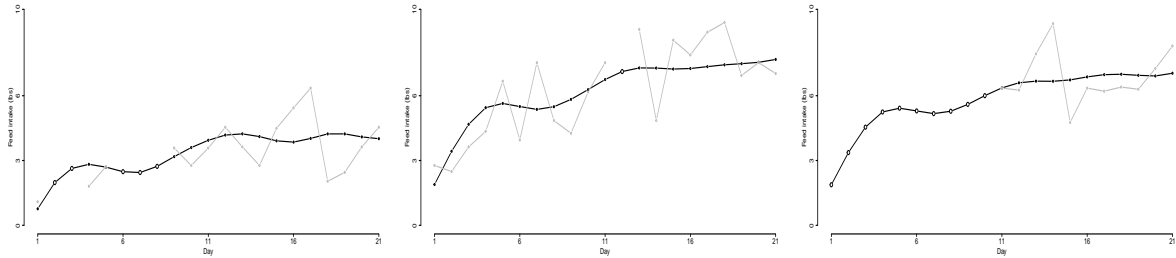


Figure 4: Prediction of full feed intake trajectories for two young sows (left and middle) and one old sow (right panel) from different groups.

## 7 Discussion

In this paper we consider longitudinal dynamic functional regression for scalar responses and functional covariates observed in a longitudinal design. We propose a flexible way to model the time-varying bivariate regression coefficient function, which is very flexible, unlike existing literature on this topic (see Kundu et al. (2016) and Goldsmith et al. (2012)) it does not rely on prior assumptions about the underlying structure. The methodology is applicable to Gaussian as well as non-Gaussian responses. It can directly accommodate additional vector covariates, non-linear effects of vector covariates, as well as multiple functional covariates observed on diverse sampling designs and with measurement error. The methodology can be easily implemented using the existing freely available software.

Numerical results show that the prediction performance of our approach is superior to existing alternative approaches when the regression coefficient function is indeed time varying, and is very competitive with the existing alternatives when the regression coefficient function is time-invariant. In spite of the increased flexibility, this method is computationally efficient; in fact it is orders of magnitude faster than its closest competitor. We discuss an approach to reconstruct the full response trajectory. We applied the method to the animal science application and found that the effect of the relative humidity on the feed intake of the lactating sows varies with the days since they gave birth.

One limitation of our methodology is that it relies on the implicit assumption that the current response is related to the current functional predictor only i.e.  $E[Y_{ij}|X_{i1}(\cdot), \dots, X_{in_i}(\cdot)] = E[Y_{ij}|X_{ij}(\cdot)]$ . While this assumption makes sense for our application, it may not be reasonable for other situations. One possible approach to account for the past functional covariates is by considering a dependence similar to the one implied by the functional linear model (see Malfait and Ramsay (2003); Pomann et al. (2016); Scheipl et al. (2015) and Kim et al. (2011)).

## Supplementary Material

Additional simulation results as well as additional data analysis results are presented as Supplementary Material. Moreover, the R-code for implementation of the proposed framework is posted publicly at <http://www4.stat.ncsu.edu/~staicu/software/LDFR.zip>. The fitting methodology is illustrated at [http://www4.stat.ncsu.edu/~staicu/software/illustration\\_LDFR.html](http://www4.stat.ncsu.edu/~staicu/software/illustration_LDFR.html) using a generated data set.

## Acknowledgment

Staicu's research was funded by National Science Foundation grant DMS 1454942 and National Institute of Health grants R01 NS085211 and R01 MH086633. The data used originated from work supported in part by the North Carolina Agricultural Foundation, Raleigh, NC.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19:716–723.
- Baumgard, L. (2015). Assessing the impact of seasonal loss of productivity. national pork board animal science webinar.
- Black, J., Mullan, B., Lorsch, M., and Giles, L. (1993). Lactation in the sow during heat stress. *Livestock production science*, 35:153–170.
- Bloemhof, S., Mathur, P., Knol, E., and Van der Waaij, E. (2013). Effect of daily environmental temperature on farrowing rate and total born in dam line sows. *Journal of animal science*, 91:2667–2679.
- Cai, T. T., Hall, P., et al. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34:2159–2179.
- Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45:11–22.
- Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13:571–591.
- Cardot, H. and Sarda, P. (2008). Varying-coefficient functional linear regression models. *Communications in Statistics Theory and Methods*, 37:3186–3203.
- Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The annals of applied statistics*, 3:458.
- Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 11:89–102.
- Eilers, P. H. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2:637–653.
- Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: a bayesian perspective. *Statistica Sinica*, 14:731–761.
- Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62:303–322.
- Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal data analysis*. CRC Press.
- Gertheiss, J., Goldsmith, J., Crainiceanu, C., and Greven, S. (2013). Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, 14:447–461.
- Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61:453–469.
- Goldsmith, J., Crainiceanu, C. M., Caffo, B. S., and Reich, D. S. (2011). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage*, 57:431–439.
- Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223.

- Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4:1022–1054.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55:757–796.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85:809–822.
- Huang, L., Scheipl, F., Goldsmith, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. (2015). *refund: Regression with Functional Data*. R package version 0.1-13.
- Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015). Penalized function-on-function regression. *Computational Statistics*, 30:539–568.
- Johnston, L., Ellis, M., Libal, G., Mayrose, V., and Weldon, W. (1999). Effect of room temperature and dietary amino acid concentration on performance of lactating sows. ncr-89 committee on swine management. *Journal of animal science*, 77:1638–1644.
- Kim, K., Şentürk, D., and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of statistical planning and inference*, 141:1554–1566.
- Kundu, M. G., Harezlak, J., and Randolph, T. W. (2016). Longitudinal functional models with structured penalties. *Statistical Modelling*, 16:114–139.
- Lawrence, M. G. (2005). The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, 86:225–233.
- Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics*, 31:115–128.
- Marx, B. D. (2010). P-spline varying coefficient models for complex data. In *Statistical Modelling and Regression Structures*, pages 19–43. Springer.
- Marx, B. D. and Eilers, P. H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, 28:193–209.
- McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23:249–269.
- Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68:179–199.
- Moyeed, R. and Diggle, P. J. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Australian Journal of Statistics*, 36:75–93.
- Müller, H.-G. (2005). Functional modelling and classification of longitudinal data\*. *Scandinavian Journal of Statistics*, 32:223–240.
- Müller, H.-G. and Yao, F. (2012). Functional additive models. *Journal of the American Statistical Association*, 103:1534–1544.
- Park, S. Y. and Staicu, A.-M. (2015). Longitudinal functional data analysis. *Stat*, 4:212–226.

- Pomann, G., Staicu, A.-M., Lobaton, E., Mejia, A., Dewey, B., Reich, D., Sweeney, E., and R, S. (2016). A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions. *Annals of Applied Statistics*.
- Ramsay, J. and Silverman, B. W. (1997). *Functional Data Analysis, 1st edition*. Springer-Verlag, Berlin.
- Ramsay, J. O. and Silverman, B. W. (2005,). *Functional Data Analysis*. Springer.
- Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102:984–996.
- Renaudeau, D. and Noblet, J. (2001). Effects of exposure to high ambient temperature and dietary protein level on sow milk production and performance of piglets. *Journal of animal science*, 79:1540–1548.
- Rosero, D. S., Boyd, R. D., McCulley, M., Odle, J., and van Heugten, E. (2016). Essential fatty acid supplementation during lactation is required to maximize the subsequent reproductive performance of the modern sow. *Animal reproduction science*, 168:151–163.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11:735–757.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Cambridge university press.
- Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24:477–501.
- St-Pierre, N., Cobanov, B., and Schnitkey, G. (2003). Economic losses from heat stress by us livestock industries. *Journal of dairy science*, 86:E52–E77.
- Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 11:177–194.
- Wood, S. (2006a). *Generalized additive models: an introduction with R*. CRC press.
- Wood, S. N. (2006b). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, 48:445–464.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36.
- Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, (just-accepted):1–45.
- Xiao, L., Li, Y., and Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75:577–599.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.
- Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, 50:689–699.
- Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95:601–619.